Strategic development in exact calculation: Group and individual differences in four achievement subtypes

Wylie, J., Jordan, J-A., & Mulhern, G. (2012). Strategic development in exact calculation: Group and individual differences in four achievement subtypes. DOI: 10.1016/j.jecp.2012.05.005

Published in:
Journal of Experimental Child Psychology

Document Version:
Publisher's PDF, also known as Version of record

Journal of Experimental Child Psychology 113 (2012) 112–130

Contents lists available at SciVerse ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Strategic development in exact calculation: Group and individual differences in four achievement subtypes

Judith Wylie *, Julie-Ann Jordan, Gerry Mulhern

School of Psychology, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, UK

A R T I C L E   I N F O

Article history:
Received 6 October 2011
Revised 9 May 2012
Available online 18 June 2012

Keywords:
Strategy choice
Mathematical difficulty
Exact calculation
Individual differences
Longitudinal
Subtypes

A B S T R A C T

This longitudinal study sought to identify developmental changes in strategy use between 5 and 7 years of age when solving exact calculation problems. Four mathematics and reading achievement subtypes were examined at four time points. Five strategies were considered: finger counting, verbal counting, delayed retrieval, automatic retrieval, and derived fact retrieval. Results provided unique insights into children's strategic development in exact calculation at this early stage. Group analysis revealed relationships between mathematical and/or reading difficulties and strategy choice, shift, and adaptiveness. Use of derived fact retrieval by 7 years of age distinguished children with mathematical difficulties from other achievement subtypes. Analysis of individual differences revealed marked heterogeneity within all subtypes, suggesting (inter alia) no marked qualitative distinction between our two mathematical difficulty subtypes.

Introduction

It is widely accepted that the mix and sophistication of strategy use in tasks involving number and arithmetic provide a useful indication of a child's developmental progress in these areas (e.g., Geary, Hoard, Nugent, & Byrd Craven, 2007; Siegler, 1996). In general, the typical development of expertise in numerical cognition is considered to involve the gradual discarding by young children of slower, less efficient reconstructive processes in favor of a mixture of more sophisticated reconstructive and retrieval-based processes and, ultimately, to the use of efficient retrieval-based strategies by older children and adults (e.g., Zbrodoff & Logan, 2005). By contrast, atypical development is characterized by

* Corresponding author. Fax: +44 28 90975486.
E-mail address: jw.wylie@qub.ac.uk (J. Wylie).

0022-0965/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jecp.2012.05.005

different mixes of strategy use and different developmental trends in the adoption of more efficient strategies, depending on the specific nature of the mathematical difficulty (Geary & Hoard, 2005). A consistent finding has been that children with mathematical difficulties show deficits in the use of retrieval-based processes, as evidenced by a persistently reduced frequency of accurate retrieval of number facts compared with typically achieving children. Moreover, these children have been found to rarely use derived fact retrieval strategies such as decomposition (Geary et al., 2009).

In typically developing children, initial attempts at calculation usually involve guessing and estimation (Jordan, Kaplan, Ramineni, & Locuniak, 2008; Jordan, Levine, & Huttenlocher, 1994). When children begin formal instruction, they are first taught to add single-digit numerosities using counting-based strategies such as finger or verbal counting, and these strategies increase in efficiency and sophistication with age (e.g., Geary et al., 2007). Eventually, after repeated practice, a long-term memory representation is formed between an answer and the problem, and there is a corresponding shift from counting- to retrieval-based procedures (Siegler, 1996; Siegler & Shrager, 1984).

Comparisons between the strategy development of typically achieving children and those with mathematical difficulties have provided important insights into the nature and underlying causes of mathematical difficulty (e.g., Geary et al., 2009; Jordan et al., 2008). Two specific aspects of strategy development are of interest, namely, strategy mix and relative effectiveness. Typically, studies have focused on the extent to which children's solution strategies are based on counting or retrieval. More specifically, researchers have tended to classify strategies, either observed or self-reported, according to the level of maturity (e.g., DeCorte & Verschaffel, 1987; Geary, Hamson, & Hoard, 2000; Hanich, Jordan, Kaplan, & Dick, 2001; Ostad, 1999; Siegler & Shrager, 1984). From less to more mature, those commonly identified have been *finger counting* or use of other physical referents, *verbal counting* not obviously involving physical referents, relatively slow *delayed retrieval* with no observable strategy, and the more mature strategies involving relatively fast *direct/automatic retrieval,* with no observable strategy and thus assumed to involve retrieval of a known fact, and *derived fact retrieval,* with the answer derived from a known fact or arithmetical principle, typically through commutativity, decomposition, associativity, or use of relationships between different operators and requiring explicit understanding of the relationships between number facts (Dowker, 2009).

Subtype differences in strategy use and effectiveness

Geary and colleagues (2000) investigated strategy choice and accuracy on simple addition problems in 7-year-olds (mean age = 82 months) and again 1 year later. They reported that at age 7, children with mathematical difficulty (MD children) and those with comorbid mathematical and reading difficulty (MDRD children) made similar numbers of counting errors. In terms of strategy mix, both subtypes exhibited approximately equal use of counting and retrieval strategies. By contrast, children with reading difficulty (RD children) and typically achieving (TA) children demonstrated virtually exclusive use of the less developmentally mature counting strategies, but with much higher accuracy levels than MD and MDRD children, suggesting that MD children were making inappropriate use of direct retrieval, whereas TA and RD children were choosing appropriately.

Geary and colleagues (2000) reported that by 8 years of age, both TA and RD children exhibited a shift toward direct retrieval, again with very low error rates, whereas MD children showed a clear shift away from retrieval toward verbal counting. MDRD children were also found to use more counting-based strategies. All four subtypes produced much lower error rates than at 7 years of age, although MDRD children still performed relatively poorly. In general, at age 8, MD children's counting accuracy has been widely reported to be greater than that of MDRD children on untimed arithmetic tasks (Geary, Hoard, & Hamson, 1999; Jordan & Hanich, 2000; Jordan & Montani, 1997) and comparable to that of TA and RD children (Geary et al., 2000; Jordan, Hanich, & Kaplan, 2003), although on forced retrieval tasks their performance is comparable to that of MDRD children.

Thus, there is strong evidence to suggest that young MD or MDRD children differ from TA children in terms of frequency and efficiency of strategy use. However, the few studies that have examined children's mathematical difficulties using longitudinal or age/ability matched designs have produced mixed evidence regarding whether strategy development in these subtypes is characterized by persistent deficits or by developmental delay. In a longitudinal study of TA children between kindergarten and second grade, Jordan and colleagues (2008) reported a general increase in the use of finger counting and a corresponding decrease in the adaptiveness of the strategy, indicated by a decrease in the correlation between frequency of use and accuracy. In other studies, MD children have been found to rarely use direct retrieval and to frequently use backup strategies, suggesting little evidence of the typical shift from immature to more mature strategies throughout the elementary school years (Geary, Brown, & Samaranayake, 1991; Ostad, 1997). This tendency has also been reported in MDRD children (Geary et al., 2000). Strategy efficiency comparisons by Geary and colleagues (1991) revealed that, over time, MD children became more skilled at finger counting but not at direct retrieval. These findings suggest that MD children may fail to make the shift from more immature counting procedures to direct retrieval due to persistent deficits in number fact retrieval.

Other researchers have suggested that fact retrieval difficulties in MD children represent a developmental delay rather than a persistent deficit. In a study of 10-year-old MD children employing an age/ability matched design (Torbeyns, Verschaffel, & Ghesquiere, 2004), MD children were found to differ from age-matched controls in terms of frequency, efficiency, and adaptiveness of strategy use. On the other hand, these children did not differ from an ability-matched group on any of these aspects, suggesting that their strategy use was characterized by a developmental delay. Furthermore, Jordan and colleagues (2003) reported that neither MD nor MDRD subtypes differed significantly from TA children in terms of overall rate of growth on a forced retrieval task, and Geary and colleagues' (2000) longitudinal study found that MD children, and to a lesser extent MDRD children, displayed considerable reductions in both procedural and retrieval errors over time.

Individual differences in strategy use and effectiveness

Although it is widely acknowledged that there are marked individual differences in components of arithmetical cognition, even among typically achieving groups of children and typical adults (Dowker, 2005), virtually all of the data presented in the literature on the development of counting and arithmetic have been at the level of the group. Recently, Jordan, Mulhern, and Wylie (2009) argued for the need to consider individual differences when studying children's arithmetical development and for adopting a longitudinal perspective to such individual differences. In a study of several components of mathematical achievement in TA children, they reported that whereas analysis of group data suggested linear growth over time, analysis of individual differences provided an altogether different picture. Children were found to differ markedly from each other on initial status, final status, growth rate, and shape of trajectory on seven arithmetical tasks and were even found to differ within themselves on these parameters across tasks. Although Jordan and colleagues studied performance rather than strategy development, their findings suggest that a focus on individual differences in strategy use over time may reveal similar findings regarding possible discrepancies between group and individual trajectories.

The issue of individual differences is compounded when achievement groups other than TA children are considered. Although previous research has revealed important findings regarding the development of arithmetical cognition within groups with singular or comorbid mathematical and reading difficulties, a consideration of the relative impact of between- and within-group effects would greatly inform the issue of the relationship between achievement difficulties and arithmetical development.

The current study

The aims of this study were to investigate the use of various strategies for exact calculation in children younger than those reported in previous research. Specifically, using a longitudinal design, we sought to identify developmental changes in strategy use by children at four time points between 5 and 7 years of age and in four achievement subtypes (TA, MD, RD, and MDRD). In addition to group differences, we investigated the extent of individual differences in strategy use over time within each subtype. We also examined the effectiveness of different strategies for exact calculation for the four subtypes. A further aim was to address whether individual differences within the MD and MDRD groups were consistent with the view that these are qualitatively different subtypes.

Method

Participants

A total of 85 children participated in the study. They had been selected from an initial screening sample of 256 children (141 boys and 115 girls) from 14 local schools representing the full range of socioeconomic status (SES) based on the Northern Ireland Multiple Deprivation Measure (Northern Ireland Statistics & Research Agency, 2005). The mean standardized nonverbal IQ for the final sample was 95.32 (*SD* = 11.79) based on the Nonverbal cluster of the British Ability Scales II (Elliott, Smith, & McCulloch, 1997).

All participating schools were part of the mainstream primary school system in Northern Ireland, which is based on the U.K. National Curriculum. Children start school in Northern Ireland from the age of 4 years 2 months, with a mean age of 4 years 8 months during the first year of schooling. Recruitment was carried out on the basis of informed written parental consent. At the time of the screening, children's mean age was approximately 5 years and all participants spoke English as their first language.

Materials and procedure

Screening measures and subtype classification

The Rhyme Detection and Phoneme Deletion (beginning sounds) subtests of the Phonological Abilities Test (PAT; Muter, Hulme, & Snowling, 1997) measured children's phonological ability. In the Rhyme Detection subtest, children selected which of three words rhymed with a stimulus word (e.g., *cat*—which word rhymes? *fish, gun*, or *hat*). For the Phoneme Deletion (beginning sounds) subtest, children orally deleted the first phoneme of a single syllable word (e.g., "bus without the [b] says . . .").

The Alphabet subtest of the Test of Early Reading Ability (TERA; Reid, Hresko, & Hammill, 2001) measures sound–letter correspondence and children's knowledge of the letters of the alphabet. More specifically, the items assess awareness of letters in different fonts and knowledge of syllables and letter names.

The Test of Early Mathematics Ability (TEMA; Ginsburg & Baroody, 2003) identifies mathematical difficulties in children 3 years 0 months to 8 years 11 months of age. It assesses a broad range of formal and informal mathematical skills, including number comparison, nonverbal arithmetic, counting, problem solving, numbering skills, numeral literacy, mastery of number facts, calculation skills, and understanding of concepts. For the purposes of the current study, a number of minor adaptations were made, for example, to items referring to American currency and names.

All tests were administered individually, with a typical testing session lasting 30 min. Children who scored at or below the 25th percentile in at least two of the three testing phases were classified as having difficulty in that particular achievement area. To be considered as typically achieving, children needed to score at or above the 40th percentile on the relevant test. The specific achievement criteria for the various subtypes were as follows: for MD, TEMA score at or below the 25th percentile and PAT and TERA Alphabet scores at or above the 40th percentile; for RD, PAT and TERA Alphabet scores at or below the 25th percentile and TEMA score at or above the 40th percentile; for MDRD, at or below the 25th percentile on all screening tests; and for TA, at or above the 40th percentile on all screening tests. Children with test scores in the 26th to 39th percentile range were unclassified.

From this screening, the most prevalent learning difficulty subtype was MD (26%), followed by MDRD (21%) and RD (19%), with 34% of the sample being typically achieving. With the exception of the TA subtype, boys outnumbered girls and, in the case of MDRD and RD, did so by a ratio of approximately 2:1. Those children allocated to the MDRD subtype were matched as closely as possible on reading ability with the RD subtype and on mathematics ability with the MD subtype. Similarly, the TA subtype was matched as closely as possible on reading ability with the MD subtype and on mathematics ability with the RD subtype. The mean percentile scores for the various subtypes are shown in Table 1. Care was taken to ensure that, as far as possible for each school, there was a similar mix of each of the four subtypes and all subtypes were matched for age.

Table 1 displays the ability information for each subtype and the number of children who received math and/or reading help throughout the longitudinal study. Although MD and MDRD children were just as likely to receive mathematics intervention, MDRD children were more likely to receive extra help with reading than were RD children.

Table 1
Subtype characteristics.

Subtype	n	Mathematics intervention (n)	Reading intervention (n)	Mathematics mean percentile	Reading mean percentile
TA	29	0	0	54.76	75.83
RD	16	0	2	50.31	21.06
MD	22	5	1	19.73	59.00
MDRD	18	3	6	19.28	11.94

Experimental task: exact calculation

The exact calculation task was based on that used by Hanich and colleagues (2001) and Jordan and colleagues (2003). Because the problems used in these studies were designed for use with children from 7 to 9 years of age, the difficulty of items was adjusted for 5- to 7-year-olds by reducing the numerical values of the operands to single digits. The number of items was also reduced from eight to six. The adaptation was piloted on a sample of children containing roughly equal numbers of participants from all subtypes. Because performance did not produce either floor or ceiling effects for any subtype, the problems were deemed to be appropriate. During the pilot phase, the three authors tested in all three combinations of experimenter pairings in order to establish the reliability of strategy identification. Discrepancies between experimenters were noted and later discussed, and this process continued until very high levels of agreement (>98%) were achieved. During the experimental phase, there was regular reliability checking and overall reliability was maintained at greater than 98% at each of the four time points.

The six single-digit arithmetic problems (three addition and three subtraction) were presented both orally and visually in horizontal orientation. Children were told that they could use whatever method they wished to get each answer and to speak the answer as soon as they knew it. Each child completed the task at four time points separated by an average of six monthly intervals, with mean ages (and standard deviations) of 65 (3.89), 71 (3.78), 77 (3.68), and 83 (3.65) months (changes in standard deviations over time reflect minor variations in sample size due to missing participants at a given time point).

Following each problem, children were asked how they had worked out the answer and the response was recorded verbatim. Based on these self-reports and on experimenter observation, each trial was classified as one of five strategies: finger counting (observed use of physical referents, invariably fingers), verbal counting (reported or observed counting without the use of fingers or other physical referents), delayed retrieval (no reported or observed strategy and response after more than 5 s, with the child assumed to be working out the answer in his or her head and relying on working memory during calculation), automatic retrieval (no reported or observed strategy and response within 5 s, assumed to involve direct recall of the solution from memory), and derived strategy use (evidence of use of strategies based on rules and relationships between number facts such as identity, commutativity, addend +1, addend −1, and addition/subtraction inverse principles) (Dowker, 1998). In practice, the distinction between delayed retrieval and automatic retrieval was very clear, with virtually all retrieval responses classified as automatic occurring within 4 s, even at Time 1, and all classified as delayed retrieval occurring beyond 5.5 s.

In cases where the experimenter's observation and the child's reported strategy were contradictory, the experimenter's observation was used if the child's method was obvious (e.g., finger counting). In cases where the experimenter was unsure which strategy the child used, the trial was classified according to the response reported by the child (Geary et al., 2000; Jordan et al., 2003).

Results

Group analysis

Overall performance by subtype

Fig. 1 presents the overall accuracy for the four achievement subtypes at each time point irrespective of strategy used. Two-way mixed analysis of variance (ANOVA) revealed no evidence of a Time $\times$ Subtype interaction but showed highly significant main effects of time, $F(3,198) = 75.93$,

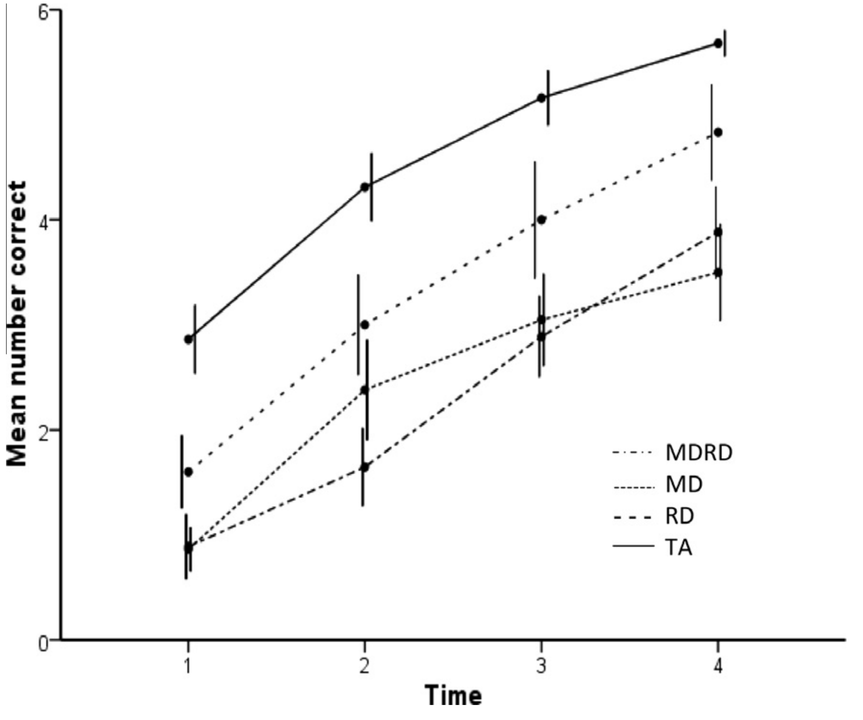

Fig. 1. Mean accuracy by subtype and time. Error bars indicate ±1 standard error of the mean.

$p < .001$, $\eta_p^2 = .48$, and subtype, $F(3,66) = 14.63$, $p < .001$, $\eta_p^2 = .41$. Polynomial contrasts revealed significant best fit linear trends across time for each subtype, and simple contrasts confirmed significant differences across all time points for all subtypes. Overall, children displayed mean accuracy rates of between 0.9 (MD/MDRD) and 2.9 (TA) items at Time 1 and between 3.5 (MD) and 5.7 (TA) items at Time 4.

Paired comparisons revealed that TA children significantly outperformed all other subtypes at all time points. RD children significantly outperformed MDRD children at Times 2 and 3 and were significantly better than MD children at Times 1 and 4. There were no significant differences between MD and MDRD children at any time point.

Strategy mix by subtype

Of the five strategies of interest, verbal counting was rarely observed throughout this developmental period for any subtype, and consequently this strategy was excluded from analysis. Overall, the two most frequently observed strategies were automatic retrieval and finger counting. At Time 1, when children were 5 years 5 months of age on average, TA and RD subtypes used the less developmentally mature strategy of finger counting most frequently (Fig. 2). MD and MDRD children, on the other hand, were found to use automatic retrieval most, followed by the two reconstructive strategies of finger counting (MD) and delayed retrieval (MDRD). By Time 4, when children were 6 years 11 months of age on average, all subtypes used automatic retrieval more frequently than finger counting.

In general, delayed retrieval was the third strategy of choice. Use of this strategy was found to decrease over time for TA children and to remain generally constant for all other subtypes. By contrast, derived fact retrieval, a more mature strategy, was not observed at all at Time 1 for any of the subtypes and virtually not at all by Time 4 for the MD and MDRD subtypes. The growth was most evident in the two subtypes without mathematical difficulty for whom derived fact retrieval became the second

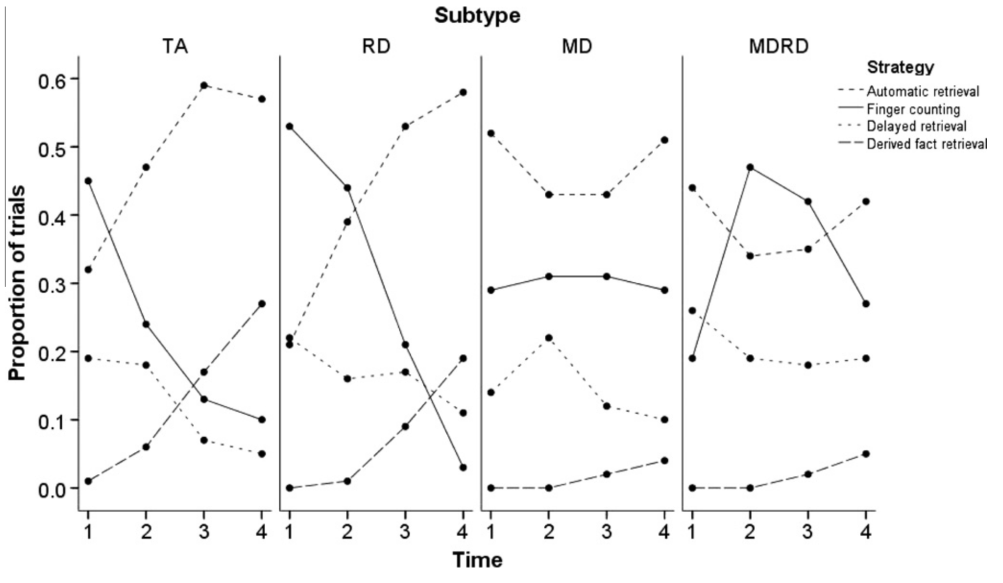

Fig. 2. Mean proportion of trials on which each strategy was used by time and subtype.

most used strategy by Time 4, overtaking finger counting (and significantly so): $t_{RD} = 2.35$, $df = 11$, $p = .04$; $t_{TA} = 2.41$, $df = 24$, $p = .02$.

Strategy shift by subtype

The key difference between the subtypes was the extent and timing of strategy shift, as revealed by a significant Time × Subtype × Strategy interaction, $F(27,585) = 2.49$, $p < .001$, $\eta_p^2 = .10$. As Fig. 2 shows, the most salient feature of this interaction was the mirror symmetry between the two most frequently used strategies, finger counting and automatic retrieval, suggesting a switch from finger counting to automatic retrieval over time, with the timing of this shift differing between subtypes. Post hoc analysis revealed that although both TA and RD children were found to favor finger counting over automatic retrieval at Time 1, the difference was significant only for the RD subtype ($p = .03$). For both math difficulty groups, there was a clear, but nonsignificant, preference for automatic retrieval at Time 1. By Time 2 (5 years 11 months of age), TA was the only subtype to use automatic retrieval significantly more often than finger counting ($p = .01$), whereas for RD children this difference reached significance ($p < .001$) only at Time 4 (6 years 11 months of age). MD and MDRD children, on the other hand, failed to demonstrate a statistically significant preference for either strategy at any time point, although by Time 4 all subtypes demonstrated the greatest use of automatic retrieval. Most notable was the more unstable nature of strategy choice by MDRD children over time compared with the other subtypes, with preference oscillating between automatic retrieval and finger counting (Fig. 2).

Subtype differences by strategy

Finger counting. Over time, there was a significant change in the use of finger counting, but this effect was modified by a Time × Subtype interaction, $F(9,198) = 3.72$, $p < .001$, $\eta_p^2 = .15$. MD children displayed virtually no change over time in the use of finger counting, whereas polynomial contrasts revealed best fit negative linear trends for RD children, $F(1,9) = 13.22$, $p = .005$, $\eta_p^2 = .60$, and TA children, $F(1,23) = 15.06$, $p = .001$, $\eta_p^2 = .40$, and a quadratic trend, $F(1,15) = 6.24$, $p = .025$, $\eta_p^2 = .29$, for MDRD children, with an initial increase in use followed by a decrease, resulting in a return to Time 1 levels by Time 4.

Delayed retrieval. There was a significant main effect of time, $F(3,198) = 3.39$, $p = .019$, $\eta_p^2 = .05$, and no Time × Subtype interaction for this relatively immature strategy, with children generally using delayed retrieval less frequently at Times 3 and 4.

Automatic retrieval. There was a significant Time × Subtype interaction, $F(9,195) = 3.87$, $p = .010$, $\eta_p^2 = .06$. A significant positive linear trend for time was apparent for the TA subtype, $F(1,23) = 10.87$, $p = .003$, $\eta_p^2 = .32$, and the RD subtype, $F(1,9) = 9.28$, $p = .014$, $\eta_p^2 = .51$. By contrast, neither the MD nor MDRD subtype showed a significant change over time, and both obtained best, but nonsignificant, quadratic fits (*p*s = .163 and .195, respectively). For both groups, the trend was characterized by an initial decrease in use between Times 1 and 2, followed by an increase between Times 3 and 4 to Time 1 levels of use.

Derived fact retrieval. There was a significant Time × Subtype interaction, $F(9,198) = 2.93$, $p = .003$, $\eta_p^2 = .13$, for this strategy. Analysis revealed significant best fit positive linear trends for the TA subtype, $F(1,23) = 32.70$, $p < .001$, $\eta_p^2 = .59$, and the RD subtype, $F(1,9) = 8.18$, $p = .019$, $\eta_p^2 = .48$, and best, but nonsignificant, linear fits for the MD and MDRD subtypes (*p*s = .131 and .122, respectively).

Effectiveness of strategies

Fig. 3 presents the effectiveness of strategy use, that is, the mean proportion correctly answered for a given strategy, as indicated by the diameters of the circles on the graphs, with diameter proportional to the number of correctly answered problems divided by the frequency of use of that strategy. Table 2 presents these effectiveness data.

Following Jordan and colleagues (2008), for the two dominant strategies of finger counting and automatic retrieval, correlations between the proportion of problems on which the strategy was used and the proportion of those trials solved correctly at each time point are presented in Table 3. Correlations were not calculated where very low levels of overall use and/or accuracy made such a calculation impractical or difficult to interpret. Specifically, at Time 4, only 1 RD child was found to use

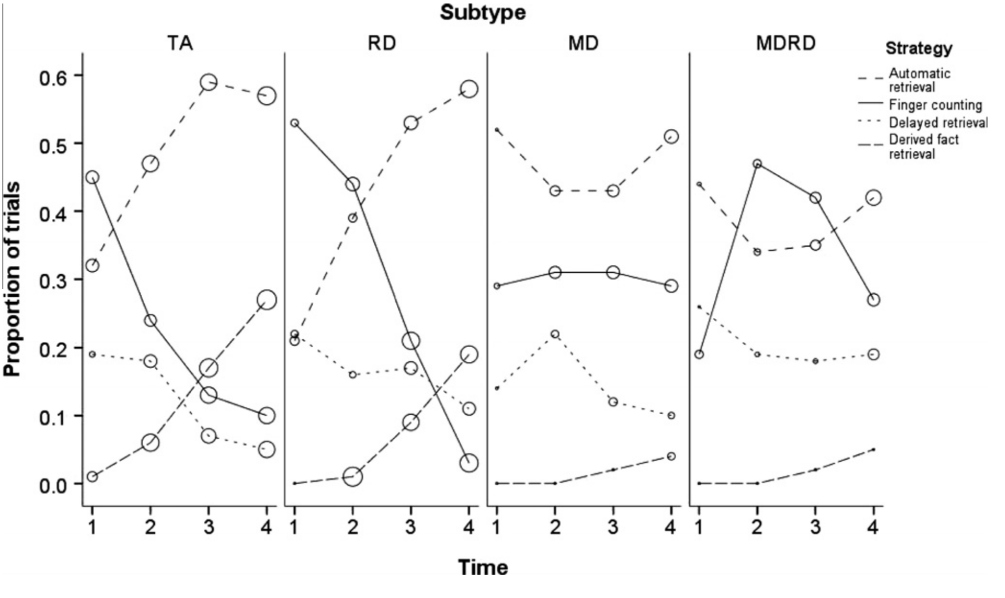

Fig. 3. Mean proportion of trials on which each strategy was used and accuracy of each strategy by time and subtype. Accuracy is indicated by circle diameters, with diameter proportional to mean proportion of trials correct.

finger counting at all, whereas at Time 1, 14 of 20 MD children and 12 of 15 MDRD children failed to obtain a single correct response using automatic retrieval. As Table 3 shows, the pattern of correlations suggests increasing adaptiveness in the use of both finger counting and automatic retrieval for all subtypes. Further statistical analysis was not feasible due to the small numbers of valid cases in each subtype arising from zero use of a given strategy by an individual child at a given time point. Instead, data are presented descriptively.

TA. Overall, this subtype showed effective use of all strategies from Time 2 onward, as indicated by the preponderance of large circles on the graph. At Time 1, TA children used automatic retrieval on approximately one third of trials, and approximately half of these were answered correctly.

The shift from finger counting to automatic retrieval is clearly evident for this group, with frequency of finger counting decreasing from approximately 0.45 of all trials to 0.1 between Times 1 and 4.

Delayed retrieval was used on approximately one quarter of trials at Time 1 but was relatively ineffective, with an average proportion of .27 of correct trials. From Time 2 onward, the frequency of use of delayed retrieval decreased consistently, although when it was used it was with markedly greater effectiveness.

Derived fact retrieval was not used at all at Time 1, but the frequency increased consistently across time to approximately one quarter of trials by Time 4. When used, this strategy proved to be highly effective.

RD. Overall, the pattern of strategy use by RD children was similar to that of the TA subtype, although the small circles at Times 1 and 2 demonstrate less accuracy for all strategies at Time 1 and for delayed retrieval and automatic retrieval at Time 2. At Times 1 and 2, accuracies of 0.43 and 0.41 of trials for automatic retrieval were in marked contrast to those of 0.67 and 0.84 for TA children.

Again, the switch from finger counting to automatic retrieval was evident. By Time 4, the use of these two strategies was very similar to that of TA children at approximately 0.1 and 0.6 of trials, respectively.

As for TA children, this subtype demonstrated a gradual increase in the use of derived fact retrieval over time, becoming the second strategy of choice by Time 4. Again, when used, this strategy proved to be highly effective, with accuracy at or exceeding 0.9 of trials.

MD. This subtype demonstrated a strikingly low accuracy at Time 1 for all strategies. Furthermore, in contrast to the TA and RD groups, automatic retrieval was used more frequently at Time 1 than finger counting but with very low accuracy. Thus, there was no evidence of switching from finger counting to automatic retrieval at Time 2, with the former used on approximately one third of trials virtually constantly across all time points. In contrast to the two non-mathematics difficulty subtypes, the pattern of use of automatic retrieval was U-shaped, with nearly total inaccuracy at Time 1 and moderate accuracy thereafter.

Frequency of delayed retrieval showed little change over time, and the strategy remained largely ineffective throughout. There was a virtually negligible increase in the use of derived retrieval, with very low levels of accuracy on the few trials where the strategy was used.

Table 2
Proportion of problems correct by subtype, time, and strategy.

Subtype	TA				RD				MD				MDRD			
Time	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Finger counting	.66	.62	.86	.85	.37	.70	.89	.95	.23	.60	.67	.67	.39	.42	.57	.62
Delayed retrieval	.26	.67	.75	.87	.33	.29	.64	.67	.11	.35	.39	.29	.02	.21	.19	.55
Automatic retrieval	.67	.84	.86	.97	.43	.41	.71	.91	.11	.48	.58	.69	.17	.29	.51	.79
Derived fact retrieval	.50	.88	.97	1.00	–	1.00	.83	.96	–	–	.00	.33	–	–	.00	.00

Table 3
Correlations between the proportion of problems on which a strategy was used and the proportion of those trials solved correctly by time point and subtype.

		Time 1	Time 2	Time 3	Time 4
TA	Finger counting	.70	.80	.94	.90
	Automatic retrieval	.64	.91	.91	.98
RD	Finger counting	.75	.86	.78	–
	Automatic retrieval	.31	.42	.66	.88
MD	Finger counting	.61	.79	.83	.71
	Automatic retrieval	–	.12	.65	.82
MDRD	Finger counting	.83	.61	.87	.93
	Automatic retrieval	–	.51	.58	.79

MDRD. This subtype demonstrated a similar pattern of strategic choice to the MD subtype for three of the four strategies. Thus, automatic retrieval produced a U-shaped pattern, albeit with a slightly lower overall level of use than the MD subtype. Unlike MD children, however, there was evidence of mirror symmetry and switching between finger counting and automatic retrieval, with automatic retrieval favored at Time 1 (albeit with very low accuracy), finger counting favored at Times 2 and 3, and much more accurate automatic retrieval at Time 4.

Overall, accuracy rates at Times 1 and 2 were low and comparable to those for MD children. Accuracy across all four time points was clearly more similar to the MD subtype than to the other two subtypes.

Frequency of delayed retrieval was slightly higher than for other subtypes and remained largely unchanged over time. Effectiveness of this strategy was very poor, albeit with some improvement at Time 4.

As for the MD subtype, MDRD children demonstrated a negligible increase in the use of derived fact retrieval with comparably low levels of accuracy.

Individual differences

Consideration of individual differences addresses two alternatives. The first is whether individuals within a subtype exhibit a signature characteristic of that subtype reflecting the group profile and distinct from other subtypes. The second is whether similar individual signatures are found across subtypes so that group patterns are an aggregate of these disparate signatures. The latter alternative may be considered to undermine the subtyping approach, suggesting the need for caution in interpreting group differences.

Fig. 4 presents an individual graph for each child, in the four subtypes at each time point, using the format adopted for group data in Fig. 3 (data for only those children who were present at all four time points are plotted). For each subtype, individual graphs are presented in order of age from youngest to oldest.

Whereas Figs. 2 and 3 showed clear group trends in use and accuracy, Fig. 4 indicates that these group trends obscure marked individual differences within all subtypes.

TA

Although the group data suggest an appropriate use of strategies at all time points, reflected in high levels of accuracy, and a shift to more mature strategies (automatic retrieval and derived fact retrieval) over time, individual graphs show that some TA children did not exhibit such patterns. Thus, for finger counting, the group frequency figure of 0.45 (Fig. 2) and accuracy of 0.66 (Table 2) at Time 1 was produced by only 9 children, 6 of whom demonstrated poor effectiveness with error rates from 0.5 (11) to 0.8 (4), 5 of whom were in the younger half of the group. Of these children, 3 (2, 5, and 16) continued to rely on finger counting at Time 3 or beyond. Conversely, 10 children showed no use, or virtually no use, of finger counting at any time point.

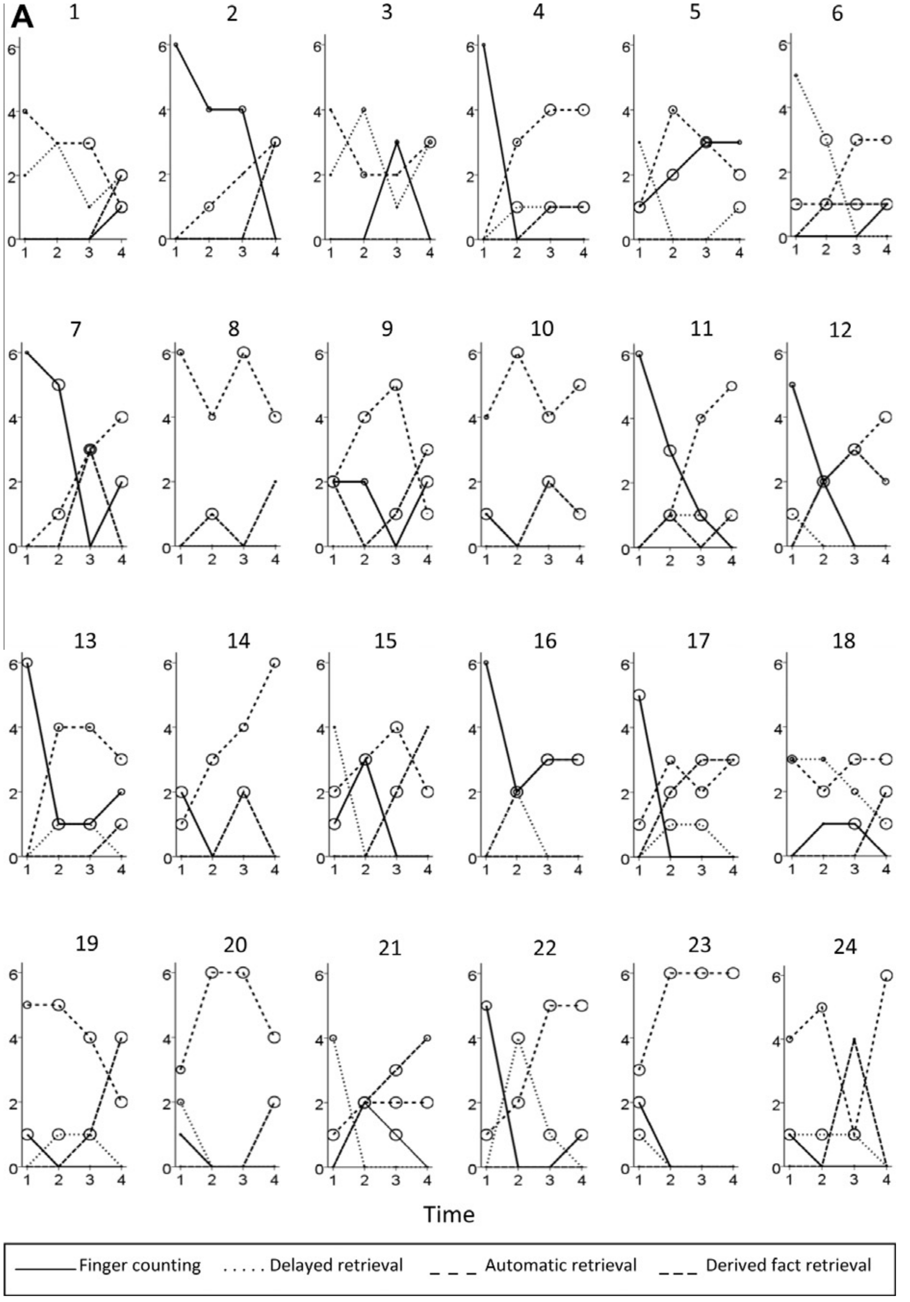

Fig. 4. (A) Individual graphs of frequency of use and accuracy by time and strategy for TA subtype. (B) Individual graphs of frequency of use and accuracy by time and strategy for RD subtype. (C) Individual graphs of frequency of use and accuracy by time and strategy for MD subtype. (D) Individual graphs of frequency of use and accuracy by time and strategy for MDRD subtype. In all panels (A–D), accuracy is indicated by circle diameters, with diameter proportional to mean proportion of trials correct.

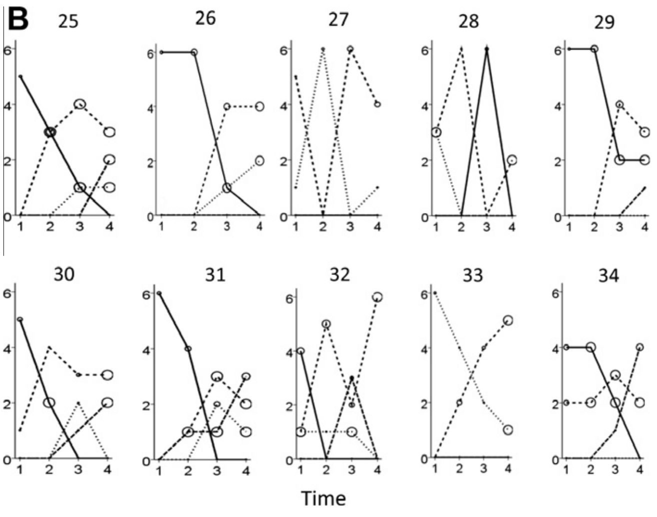

Fig. 4 (*continued*)

Similar individual differences were evident for delayed retrieval, with heavy, and largely ineffective, reliance on this strategy by 6 children but virtually no use from Time 2 onward by approximately 80% of children.

Overall, approximately half of TA children failed to exhibit the group trend for use of automatic retrieval seen in Figs. 2 and 3. Although the majority demonstrated little or no use of automatic retrieval at Time 1, by Time 4 the use of this strategy by TA children proved to be highly effective, with the exception of two younger children (1 and 9).

There was a clear trend of increasing and highly effective use over time of derived fact retrieval for TA children, albeit to a group average proportion of 0.25 of trials. However, the group trend obscures marked individual differences, with 8 children showing no use of derived fact retrieval at Time 4, 5 of whom did not use the strategy at any time point.

RD

Across all four strategies, the individual graphs reveal particular difficulties for 3 RD children (27, 28, and 30), with high error rates on all problems at all or most time points.

Fig. 4B reveals broadly similar patterns to the TA subtype. For finger counting, of the 10 children, 2 demonstrated no use of the strategy at all, and of the others, 6 did so ineffectively. Dysfunctional use of delayed retrieval was evident for 2 RD children (27 and 33). Indeed, these children exhibited exclusive reliance on delayed or automatic retrieval, with consistently high error rates at Times 1 to 3. Overall, the group error rate for automatic retrieval at Time 1 was due largely to 1 child (27), whereas the relatively high group accuracy at Times 3 and 4 (Table 2) obscured persistent ineffective use by this child.

For derived fact retrieval, individual graphs suggest that the group effect of a gradual increase in frequency of use was due to only half of the children, with 5 exhibiting no use whatsoever or use on only one trial at any of the four time points.

MD

Across all four strategies, individual graphs reveal particular difficulties for approximately half of the MD children, with high error rates on all problems at all time points, with most of these children being among the youngest in the group.

Of the 20 children, 5 relied almost exclusively on finger counting at Time 1, in all cases demonstrating high or total inaccuracy. One child (50) continued to show highly ineffective use of finger counting across the four time points, whereas another child (40) used the strategy at Times 1 and 4 with vir-

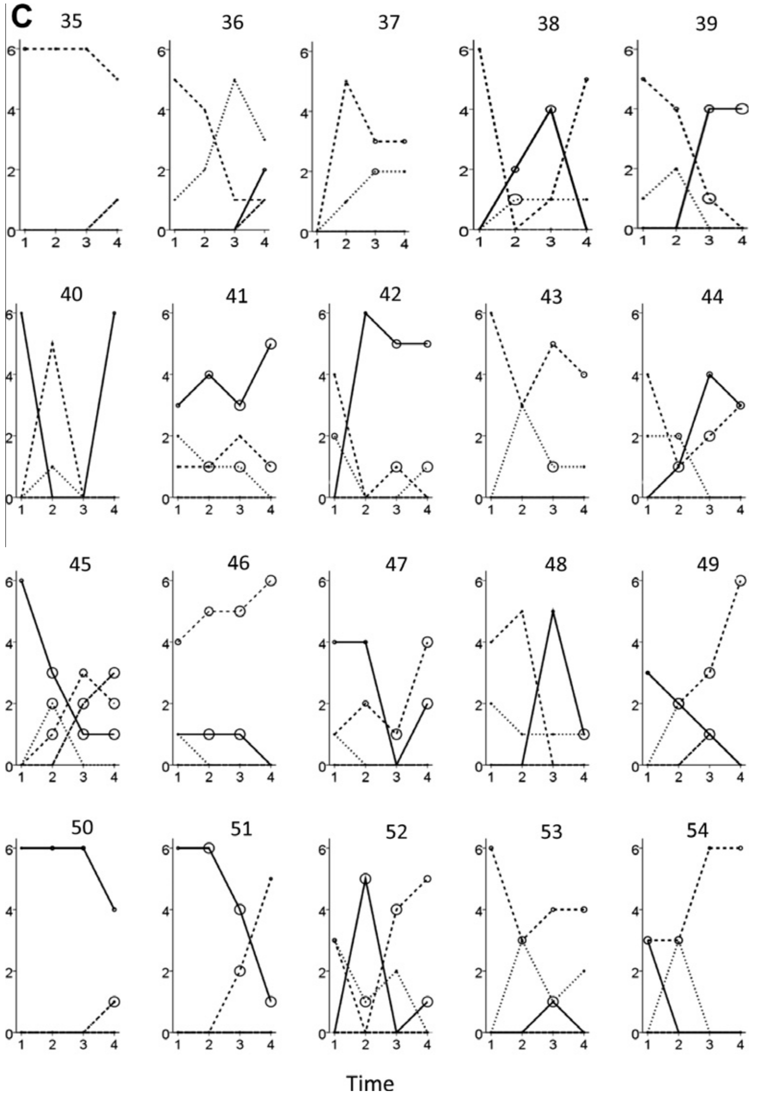

Fig. 4 (*continued*)

tually total inaccuracy. In addition, 3 children did not use finger counting at any time, and 6 used delayed retrieval virtually not at all. In general, the use of delayed retrieval was almost always dysfunctional for all MD children.

The U-shaped group trend in frequency of use of automatic retrieval over time seen in Figs. 2 and 3 was not evident in the individual graphs, suggesting little or no generalized group tendency. Individual graphs revealed interesting trends in automatic retrieval, with half of the children using the strategy at Time 1 with virtually complete inaccuracy, a tendency also evident in many MD children at Time 2. The youngest child in the group used automatic retrieval virtually exclusively and completely ineffectively at all four time points. Among the older half of the group, most were found to use the strategy effectively, particularly from Time 2 onward.

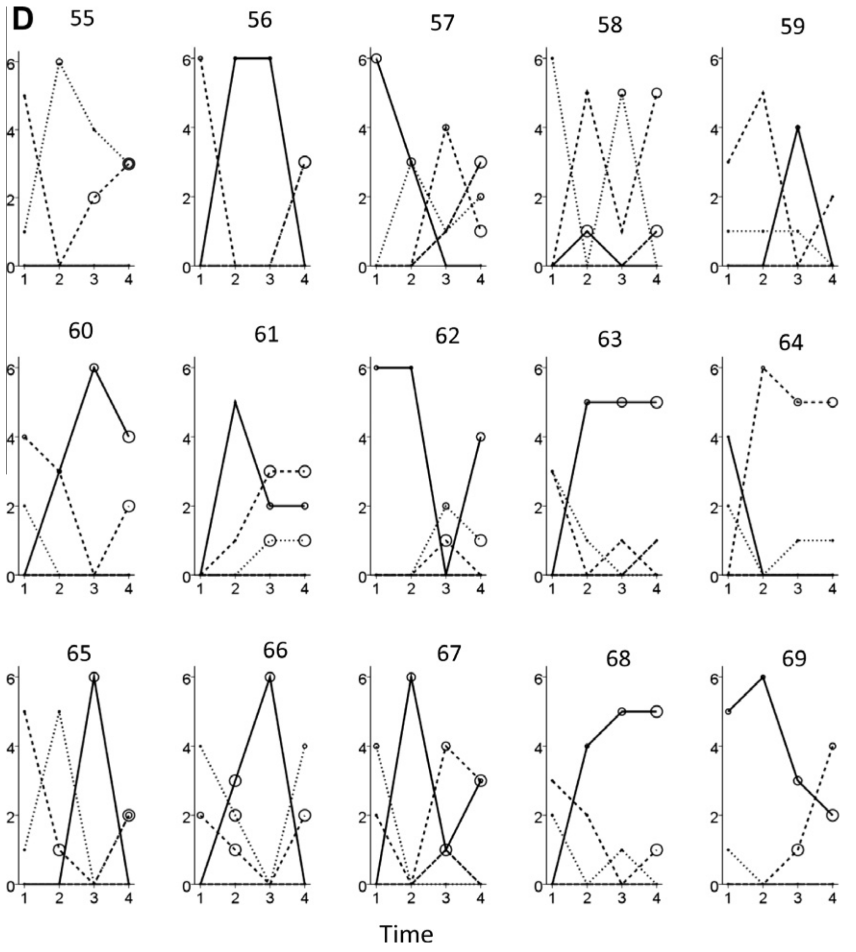

Fig. 4 (*continued*)

For derived fact retrieval, a marked difference from TA and RD children is evident, with all but 1 child (45) showing no use, or virtually no use, of the strategy at any time point. Thus, the group trend of a very shallow increase in frequency of use was due solely to this child.

MDRD

For all four strategies, individual graphs revealed particular difficulties for 4 MDRD children, with high error rates at all time points. Most did not use finger counting at all at Time 1, and the inverted U group trend in use over time (Fig. 3) is due to fewer than half of the children. A higher proportion of this subtype was found to use delayed retrieval to some degree compared with other subtypes, and the relative ineffectiveness of the strategy (Table 2) was evident in virtually all participants.

For automatic retrieval, at Time 1, a similar tendency to the MD subtype was found, with 6 of the participants using the strategy completely ineffectively. The strategy was found to be more effective at later time points for approximately half of the children. As for the MD subtype, the effective use of derived fact retrieval was confined to 1 child (57), with very scant use by 3 other children. For the MD and MDRD subtypes, the children who used this strategy did so at later time points, in contrast to TA children whose use was spread more evenly across time.

Discussion

The current study focused on the issues of strategy use and effectiveness in exact calculation, investigating whether there was evidence of differences among the four achievement subtypes at a younger age than previously reported. To date, the study is unique in considering individual differences in patterns of strategy use and their effectiveness. A supplementary aim was to inform the debate regarding whether MD and MDRD may be considered to be qualitatively distinct subtypes.

Overall group trends in accuracy

We found that the overall performance over time, aggregated across strategies, supported previous findings of subtype differences. For all subtypes, performance on exact calculation improved significantly between each time point. Consistent with previous studies involving older samples (Hanich et al., 2001; Jordan & Hanich, 2000; Jordan et al., 2003), TA children significantly outperformed all other subtypes. RD children performed better than both mathematical difficulty subtypes at all time points, although not always at a level of statistical significance.

Overall accuracy at Time 1 was low for all subtypes and especially for the two mathematical difficulty groups. By Time 4, the performance of both MD and MDRD children was only marginally above that of the TA group at Time 1, suggesting a developmental lag of approximately 18 months in exact calculation. RD children's performance suggests a lag of approximately 6 months behind TA children.

Strategy use

At Time 1 (mean age = 5 years 5 months), finger counting was found to be the dominant strategy for the TA and RD subtypes, whereas automatic retrieval was dominant for both math difficulty subtypes. By Time 4, automatic retrieval was the more frequently used and more accurate strategy for all subtypes.

To varying extents, all subtypes exhibited a shift from the use of less to more mature strategies previously reported to be characteristic of TA children (e.g., Geary et al., 2000; Hanich et al., 2001; Siegler, 1996). Notably, however, subtypes differed in terms of the timing and magnitude of this change.

The shift away from the less mature strategy of finger counting was evident after Time 1 in TA children and Time 2 in RD children. A gradual decline in the use of delayed retrieval in both groups was also evident from Time 1. For MD and MDRD children, use of delayed retrieval remained fairly constant across the four time points, albeit for approximately only one fifth of problems. Both subtypes also showed higher levels of use of finger counting relative to TA and RD children at Times 3 and 4. Finger counting in MDRD children also showed a distinctive inverted U shape, with a sharp increase from an arguably inappropriately low level at Time 1 to a peak at Time 2, little change at Time 3, and a sharp decrease at Time 4, albeit to a level similar to that at Time 1. For all subtypes, there was clear mirror symmetry in use of the two more dominant strategies of finger counting and automatic retrieval. MD children displayed stable use of finger counting and delayed retrieval over time, a finding consistent with previous research suggesting that the strategy development of children with mathematical difficulties is characterized by more persistent use of immature strategies (Geary et al., 1991, 2000; Ostad, 1997). MDRD children also showed a high degree of stability of delayed retrieval over the four time points.

By Time 4, there had been a marked shift from less to more mature strategies for both TA and RD children, with finger counting and delayed retrieval at very low levels of use. A striking feature of our findings was the progressively decreasing slopes associated with derived fact retrieval from TA through MDRD children. By Time 4, derived fact retrieval was the second preferred strategy for both TA and RD subtypes. By contrast, for both mathematical difficulty subtypes, finger counting and delayed retrieval remained much preferred strategies over automatic retrieval and derived fact retrieval across all time points. At the level of group analysis, the preference for finger counting over more mature strategies may be a distinguishing feature of mathematical difficulty between 5 and 7 years of age.

At Time 1, both math difficulty subtypes demonstrated a marked preference for the use of automatic retrieval but with very low success rates, suggesting inappropriate strategy choice at 5 years 5 months of age. Both groups were found to make inappropriate use of this more developmentally mature strategy on approximately half of all trials. By contrast, TA and RD children used automatic retrieval on only one third of trials but with much greater accuracy.

Our findings regarding the use and effectiveness of finger counting by TA children contrast with those of Jordan and colleagues (2008), who reported increasing use of finger counting from approximately 10% of problems at a mean age of 5 years 7 months to 45% of trials at around a mean age of 7 years 0 months. We found a sharp decline in the use of finger counting by TA children across this age range, a comparable decline in RD children, and no decline in MD children. Only our MDRD children showed an increase in the use of finger counting; however, this trend was reversed between 6 and 7 years of age.

In terms of effectiveness, whereas Jordan and colleagues (2008) reported a gradual decrease in the correlation between use and accuracy for TA children over time, we found this correlation to increase markedly between Times 1 and 4, with similar trends for RD, MD, and MDRD subtypes. Our results suggested that for all subtypes, the use of finger counting became more adaptive over time, although the overall frequency of use declined. For children with math difficulties, our findings support those of Geary and colleagues (1991) that, over time, MD children became more skilled at finger counting.

For the other dominant strategy, automatic retrieval, we found highly adaptive use by TA children across all time points. A similar trend was evident for RD children from Time 3 onward, but more patchy trends were found for the two mathematical difficulty groups, although by Time 4 these children also showed adaptive use of automatic retrieval. Geary and colleagues (1991, 2000) reported that MD children did not become more skilled at direct retrieval of number facts, suggesting persistent deficits in number fact retrieval. Our data do not support this contention; rather, they suggest a developmental delay in adaptive use of automatic retrieval of 12 months in RD children and approximately 18 months in the MD and MDRD subtypes.

Individual differences

As previously reported by Jordan and colleagues (2009), group trends were found to obscure important individual differences within subtypes. There was considerable heterogeneity within all four subtypes in terms of both use and effectiveness of the four strategies across time. Overall, TA children were highly adaptive in their use of strategies. However, at Time 1, more than one third of children used finger counting virtually exclusively with mixed success. Of the children who did so ineffectively, it is notable that all but 1 child was among the younger half of the TA group and, thus, may have been attempting to compensate for relatively weak number fact knowledge. The 3 children who used finger counting effectively at Time 1 were older and presumably made fewer counting errors, suggesting more effective working memory processing. From Time 2 onward, finger counting dropped off rapidly in most TA children, although approximately 10% continued to use the strategy relatively frequently at Times 3 and 4 but with more success.

The pattern of finger counting among the TA subtype offers partial support to Jordan and colleagues' (2008) finding that children did not engage in finger counting at an early stage inasmuch that this was so for approximately half of our TA sample at Time 1 but emphatically not so for another one third. Similar patterns of use were evident for the RD and MD subtypes, whereas MDRD children exhibited even less use of finger counting at Time 1 with a correspondingly greater, albeit highly inaccurate, use of automatic retrieval.

One of the more striking findings was the pattern of use and effectiveness of derived fact retrieval by the four subtypes. Consistent with the group analysis, individual patterns of use appeared to distinguish the two mathematical difficulty groups from TA and RD children. Although the overall use of the strategy was low, there was evidence of some use by most TA and approximately half of RD children, compared with only 20% of MD and 27% of MDRD children. Such patterns of use and effectiveness suggest that the mature strategies of automatic retrieval and derived fact retrieval are closely related and that derived fact strategies are not available to children with limited access to known facts.

Our findings support previous observations (e.g., Canobi, Reeve, & Pattison, 1998; Dowker, 2005) that children who are reasonably good at calculation are more likely to use derived fact strategies.

Delayed retrieval appeared to distinguish the math difficulty subtypes from the other two subtypes, with virtually all MD and MDRD children using the strategy to some degree and doing so largely ineffectively. By contrast, there was evidence of more effective use of delayed retrieval in the RD and TA subtypes. The ineffective use by the vast majority of math difficulty subtypes suggests that these children resorted to delayed retrieval for problems of which they had poor knowledge and for which counting or slow retrieval resulted in errors.

The very high and accurate use of automatic retrieval, especially from Time 2 onward, distinguished the TA children from the other subtypes. However, even in this group, several children showed relatively high levels of inaccuracy at Times 1 and 2. As with delayed retrieval, this strategy appeared to distinguish the math difficulty subtypes from the RD and TA children, with higher levels of error or virtually total avoidance of automatic retrieval.

The relatively large number of schools from which the children were recruited, and associated differences as a result of teacher effects and socioeconomic status, might be considered as candidate explanations for such heterogeneity. We contend, however, that these factors offer little explanation for the effects observed. As described in the Method section, socioeconomic status was scored at the school level (individual indicators were not available) and care was taken during the screening and final sample selection stages to ensure a range comparable to that of the general population. Regarding teaching and support differences, all schools taught the same curriculum and, as shown in Table 1, relatively low levels of support for reading and math difficulties were provided in all subtypes. Rather, we suggest that the heterogeneity of strategy use and effectiveness within the four subtypes reflects fundamental individual differences in the development of arithmetic, an understanding of which, as previously proposed by Dowker (2005), can be developed only via a longitudinal, multidisciplinary research effort.

Are MD and MDRD qualitatively different subtypes?

Both the MD and MDRD subtypes initially showed high preference for, and ineffective use of, automatic retrieval, which was in distinct contrast to the TA and RD subtypes. Thereafter, the two math difficulty groups differed only in terms of their use of finger counting, which was highly consistent and effective among MD children but less consistent and effective among MDRD children. That said, from Time 2 onward, and particularly when individual differences within groups were considered, the MD and MDRD children shared many common characteristics, particularly in terms of modest growth in the use of derived fact retrieval, little change in automatic retrieval, and maintenance of relatively high use of finger counting over time compared with TA and RD children.

On the whole, our results do not provide evidence that the MD and MDRD subtypes are qualitatively distinct groups. In fact, the overlap among all subtypes is evident from the individual graphs (Fig. 4), which show how group averages have been achieved. Such overlap would help to explain the relative instability of subtype classifications over time reported in the literature (e.g., Geary, 1990; Geary et al., 2000; Jordan, Wylie, & Mulhern, 2010; Jordan et al., 2003).

Several factors may account for differences between findings in various studies, including the age at which classification took place (in our case ~5 years) and the screening tests used. Issues around the factors underlying mathematical difficulty may also be relevant and may affect classification. There is uncertainty over the types of deficit that may contribute to children's mathematical difficulties (Swanson & Kim, 2007) and over the extent to which these may occur in isolation or may co-occur in various combinations. To date, numerous deficits have been linked to the MD subtype, including poor number sense (Butterworth, 1999), visuospatial difficulties (Rourke & Conway, 1997), and executive dysfunction (Geary et al., 2007). It is unclear as to which areas of strategy development these may affect and whether certain deficits are more detrimental than others. If multiple cognitive deficits do underlie children's mathematical difficulties, then it might not be surprising if there were greater variation within than between the MD and MDRD subtypes. Furthermore, as a consequence of differences in methods of classification, there may be variation between studies in terms of the patterns of cognitive deficit that underlie the children's mathematical difficulties in a given sample.

Conclusions

Our study provides some unique insights into the strategic development of children younger than those studied previously. We found differences between RD children and both mathematical difficulty subtypes. At 5 years 5 months of age, TA children were found to differ from other subtypes in the effectiveness of their strategy choices, although the difference between TA and RD children was less apparent, especially at Time 2 (5 years 11 months of age).

Our longitudinal data concerning typical development of finger counting contrast with those of Jordan and colleagues' (2008) sample of 7- to 9-year-olds in terms of both frequency of use and adaptiveness.

Notwithstanding our observations of heterogeneity within all subtypes, MDRD children may be characterized by poor strategy choice at 5 years 5 months of age, favoring automatic retrieval, some reliance on finger counting, and delayed retrieval at 6 years 11 months of age, and almost no use of derived fact retrieval. MD children were also found to strongly favor automatic retrieval and finger counting, with relatively little change between 5 years 5 months and 6 years 11 months of age. MD children also exhibited virtually no use of derived fact retrieval.

RD children showed greater similarity to the TA group than to either the MD or MDRD group, with very similar patterns of change over time in use of finger counting and automatic retrieval. At both group and individual levels, these children showed notable growth in the use of derived fact retrieval. Arguably, the emergence or not of this strategy between 5 years 5 months and 6 years 11 months of age is one of the more striking contrasts between children with mathematical difficulty and those without.

References

Butterworth, B. (1999). *The mathematical brain*. London: Macmillan.

Canobi, K. H., Reeve, R. A., & Pattison, P. E. (1998). The role of conceptual understanding in children's addition problem solving. *Developmental Psychology, 34*, 882–891.

DeCorte, E., & Verschaffel, L. (1987). The effect of semantic structure on first graders' strategies for solving addition and subtraction word problems. *Journal for Research in Mathematics Education, 18*, 363–381.

Dowker, A. (1998). Individual differences in arithmetical development. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 275–302). London: Taylor & Francis.

Dowker, A. (2005). *Individual differences in arithmetic*. Hove, UK: Psychology Press.

Dowker, A. (2009). Use of derived fact strategies by children with mathematical difficulties. *Cognitive Development, 24*, 401–410.

Elliott, C., Smith, P., & McCulloch, K. (1997). *British ability scales* (2nd ed.). London: NFER–Nelson.

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology, 49*, 363–383.

Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., & Nugent, L. (2009). First-grade predictors of mathematical learning disability: A latent class trajectory analysis. *Cognitive Development, 24*, 411–429.

Geary, D. C., Brown, S., & Samaranayake, V. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology, 27*, 787–797.

Geary, D. C., Hamson, C., & Hoard, M. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology, 77*, 236–263.

Geary, D. C., & Hoard, M. K. (2005). Learning disabilities in arithmetic and mathematics: Theoretical and empirical perspectives. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 253–267). New York: Psychology Press.

Geary, D. C., Hoard, M., & Hamson, C. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology, 74*, 213–239.

Geary, D. C., Hoard, M., Nugent, L., & Byrd Craven, J. (2007). Strategy use, long-term memory, and working memory capacity. In D. Berch & M. Mazzocco (Eds.), *Why is math so hard for some children?* (pp 83–105). Baltimore, MD: Paul H. Brookes.

Ginsburg, H., & Baroody, A. (2003). *Test of early mathematics ability–Third edition*. Austin, TX: Pro-Ed.

Hanich, L., Jordan, N., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology, 93*, 615–626.

Jordan, N., & Hanich, L. (2000). Mathematical thinking in second-grade children with different types of learning difficulties. *Journal of Learning Disabilities, 33*, 567–578.

Jordan, N., Hanich, L., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with co-morbid mathematics and reading difficulties. *Child Development, 74*, 834–850.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2008). Development of number combination skill in the early school years: When do fingers help? *Developmental Science, 11*, 662–668.

Jordan, N., Levine, S., & Huttenlocher, J. (1994). Development of calculation abilities in middle and low income children after formal instruction in school. *Journal of Applied Developmental Psychology, 15*, 223–240.

Jordan, N., & Montani, T. (1997). Cognitive arithmetic and problem solving: A comparison of children with specific and general mathematics difficulties. *Journal of Learning Disabilities, 30*, 624–634.

Jordan, J. A., Mulhern, G., & Wylie, J. (2009). Individual differences in trajectories of arithmetical development in typically achieving 5- to 7-year-olds. *Journal of Experimental Child Psychology, 103*, 455–468.
Jordan, J. A., Wylie, J., & Mulhern, G. (2010). Phonological awareness and mathematical difficulty: A longitudinal perspective. *British Journal of Developmental Psychology, 28*, 89–107.
Muter, V., Hulme, C., & Snowling, M. (1997). *Phonological abilities test*. London: Psychological Corporation.
Northern Ireland Statistics and Research Agency (2005). *Northern Ireland multiple deprivation measure, 2005*. Belfast, UK: Author.
Ostad, S. (1997). Developmental differences in addition strategies: A comparison of mathematically disabled and mathematically normal children. *British Journal of Educational Psychology, 67*, 345–357.
Ostad, S. (1999). Developmental progression of subtraction strategies: A comparison of mathematically normal and mathematically disabled children. *European Journal of Special Needs Education, 14*, 21–36.
Reid, D. K., Hresko, W. P., & Hammill, D. D. (2001). *Test of early reading ability 3*. Austin, TX: Pro-Ed.
Rourke, B. P., & Conway, J. A. (1997). Disabilities of arithmetic and mathematical reasoning: Perspectives from neurology and neuropsychology. *Journal of Learning Disabilities, 30*, 34–46.
Siegler, R. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.
Siegler, R., & Shrager, J. (1984). Strategy choice in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills*. Hillsdale, NJ: Lawrence Erlbaum.
Swanson, L., & Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence, 35*, 151–168.
Torbeyns, J., Verschaffel, L., & Ghesquiere, P. (2004). Strategy development in children with mathematical disabilities: Insights from the choice/no-choice method and the chronological-age/ability-level-match design. *Journal of Learning Disabilities, 37*, 119–131.
Zbrodoff, N. J., & Logan, G. D. (2005). What everyone finds: The problem size effect. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 331–345). New York: Psychology Press.